DOCUMENT RESUME

ED 319 812 TM 015 133

AUTHOR Horgan, Dianne
TITLE Students' Predictions of Test Grades: Calibration and Metacognition.
PUB DATE Apr 90
NOTE 6p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Attribution Theory; Cognitive Processes; *College Students; Feedback; *Grade Prediction; Higher Education; *Metacognition; Multiple Choice Tests; *Self Evaluation (Individuals); *Student Attitudes; *Study Habits
IDENTIFIERS *Calibration

ABSTRACT

Calibration, defined as the accuracy with which students can predict their own performance, was studied. The role of feedback in the development of calibration was examined, and questions of whether better calibration was associated with better performance and more efficient study were explored. Subjects were 136 college students in an introductory educational psychology class. Before each of four examinations, students were asked what percentage of the multiple-choice questions they expected to answer correctly and how long they had studied. After each examination, the students were asked how satisfied they were with their performance, the extent to which they attributed performance to stable or temporary causes, and how much they planned to study for the next test. Students lowered their expectations somewhat over the semester, although they maintained about the same level of satisfaction and stability of attributions. They planned to study slightly less as the semester progressed. Feedback did improve calibration. Among students of the same age making the same course grade, those who were well calibrated had better grades on unit examinations. Better calibration was associated with more efficient study habits. Four tables present the results of the study. (SLD)

# STUDENTS' PREDICTIONS OF TEST GRADES: Calibration & Metacognition

Dianne Horgan
Memphis State University

**Students' Predictions of Test Grades: Calibration and Metacognition**

Dianne Horgan
Foundations of Education
Memphis State University
Memphis, TN 38152

AERA, April 1990, Boston

**Objectives.** Despite the rapid growth of research on metacognition, one metacognitive phenomenon has been studied very little: students' ability to predict their own performance on a task. What students think they can do, rather than what they can actually do, often crucially determines their goal choices, their motivational levels in performing tasks and pursuing objectives, and their post-performance self-evaluations. This study focuses on calibration, i.e., the accuracy with which students can predict their own performance. The role of feedback in the development of calibration is examined, and also the question of whether better calibration is associated with better performance and more efficient study.

**Theoretical framework**. Current literature offers conflicting views about how well people can predict their performance. The general conclusion from the calibration literature is that most people confronted with most tasks are poorly calibrated (Lichtenstein, Fischhoff, & Phillips 1982). Glenberg & Epstein (1987), for example, reported that college students are poorly calibrated with regard to the state of their knowledge prior to tests: some students are overconfident and think they know more than they do; others are underconfident and study much more than necessary. On the other hand, literature in metacognition and self-assessment is more optimistic (e.g., Yussen, 1985; Mabe & West, 1982), showing that children become better able to monitor their own cognitive abilities with age and hence, and presumably better calibrated. In the self-assessment literature, there are typically positive correlations between self and supervisors' assessments of performance.

Despite obvious similarities in research questions, the literatures in calibration, metacognition, and self assessment have rarely overlapped. Most metacognitive research has focused on metamemory and metacomprehension; other areas of performance have been mostly unexplored. Little systematic work has been done on the *accuracy* or the *process* of self assessment or prediction. One reason for the different outcomes is that calibration researchers ask their questions in terms of probability estimates while researchers in self assessment and metacognition treat self evaluation in a broader perspective: self assessment research asks whether self assessors are any worse than other assessors, while metacognition research assumes that self assessment is a necessary component of metamemory or metacomprehension and have rarely examined it as a separate process. Hence, the methods used to measure calibration are very different from those typically used in metacognitive or self assessment studies.

There are thus two significant shortcomings. First, as Vertinksy (1986) has stated, calibration studies use artificial problems and general knowledge questions (e.g, "how sure are you that you know the capital of Maine?") rather than naturalistic tasks. Second, metacognitive and self assessmer search has rarely examined the accuracy of predictions.

This study, then, integrates these areas of research. Students are presented with a naturalistic task of predicting test performance, and the accuracy of their predictions is evaluated. It was hypothesized that as the semester progressed, predictions of performance would improve. While Glenberg & Epstein (1987) found poor calibration, their study was not done in an actual course and did not

give the students feedback from several tests. Artificial situations may underestimate calibration; Keren (1987) has argued that calibration only occurs when learners care about their performance. A second hypothesis was that, other things being equal, good calibration would be associated with better performance and more efficient study.

**Data source**. 136 college students enrolled in an introductory educational psychology class participated. Mean age was 22, with a range of 18 to 59; 91% were female.

**Methods**. Before each of four exams, Ss were asked what percentage of questions (multiple choice) they expected to get correct and how long they had studied. After each exam, they were asked how satisfied or dissatisfied they were with their performance, the extent to which they attributed their performance to stable or temporary causes and whether they planned to study more, less, or about the same on the next exam. Age, gender, and actual test scores were available for all Ss.

**Results**. Table 1 shows that students lowered their expectations slightly as the semester progressed, maintained about the same level of stability of attributions and satisfaction, and planned to study slightly less as the semester progressed.

Table 1. Means for selected variables by tests

| | Actual scores | expected | hrs studied | Satisfaction[a] | Stability[b] | Study more?[c] |
|---|---|---|---|---|---|---|
| 1 | 62.87 | 79.53 | 4.2 | 3.17 | 3.90 | 1.10 |
| 2 | 65.19 | 77.10 | 4.9 | 2.77 | 3.17 | 1.20 |
| 3 | 57.84 | 76.71 | 4.5 | 3.06 | 3.27 | 1.24 |
| 4 | 66.05 | 75.38 | 4.3 | 2.88 | 2.86 | 1.31 |

[a]1=very satisfied; 4=very dissatisfied

[b]1=very stable; 5=very temporary

[c]1=study more; 2=study about the same; 3=study less

Does calibration develop with feedback? Correlations between actual and expected test scores increased over the semester (rs=.10, .17, .43 ($p<.01$), and .43 ($p<.01$) for tests 1 through 4, respectively). Since older students studied more and made better grades (average rs=.335 and .278), the partial correlations, with the effects of age removed, were also computed. Partial correlations were comparable, although slightly lower. This pattern shows that calibration did improve with feedback.

Is calibration associated with better performance? Accuracy of calibration can be computed by the following formula (Lichtenstein & Fischhoff, 1977):

$$\frac{(\text{actual score} - \text{expected score})^2}{\text{expected score}}$$

To determine whether there is a meaningful relationship between accuracy of prediction on a specific test and performance on that test, it is necessary to control for overall ability within the domain. Glenberg & Epstein (1987) argue that students predict their performance based on a global self evaluation--"I am an A student in physics, therefore I should make an A on this physics exam." Although their predictions will often be correct, they may not know what

knowledge they have for a particular exam. Further, they may not study efficiently--they may have studied more than necessary. Hence, their calibration could be poor. The best measure available of overall ability in the domain of educational psychology is the course grade. The four unit exams accounted for 60% of the course grade, 20% was from a final exam, and the remaining 20% from written work. Since age relates to test performance, this variable also needs to be controlled. To control for the quality and the age of the student , overall course grade, age, and accuracy (as measured by the above formula) were regressed onto actual test performance. The resulting partial correlations between accuracy of calibration and actual performance on all four exams, holding both age and overall performance constant, are shown in Table 3 and are all statistically significant ($p<.01$). These results show that among students of the same age who made the same course grade, those who were well calibrated made better grades on unit exams. Students who "know what they know" *for a particular exam* perform better on that exam.

Table 3. Partial correlations between actual test scores and calibration accuracy

| | pr(Accuracy by actual test score)[a] |
|---|---|
| Test 1 | -.320** |
| Test 2 | -.208** |
| Test 3 | -.439** |
| Test 4 | -.278** |

**p<.001

[a]with the effects of age and course grade removed

Is calibration associated with more efficient study? Better calibrated students ought not to overstudy. The relationship between study time and accuracy of calibration is difficult to examine since well calibrated students tend to make better grades, which is, in general, associated with studying *more*. (average rs=.196). Table 4 shows the partial correlations between the number of hours studied for each exam and the accuracy of calibration for that exam, holding age and actual test scores constant.

Table 4. Partial correlations hours studied and calibration

| | pr(accuracy by hours)[a] |
|---|---|
| Test 1 | .161 |
| Test 2 | -.123 |
| Test 3 | .363** |
| Test 4 | .266* |

**p<.001

[a]with the effects of age and actual grade removed

By test 3, among students of the same age who scored the same on the exam, those who were well calibrated studied *less*. In other words, while in general, studying more leads to higher grades, being well calibrated allows one to know when to stop studying.

**Conclusion.** In contrast to typical calibration findings, these results on a naturalistic task, showed that students improved predictions over the semester. Better calibration was associated with better performance and more efficient study habits.

**Educational and scientific importance.** Motivation depends upon *subjective* expectations of success; students try to do what they *think* they can do. Underconfidence can lead to lower performance--students do not try because they expect to fail. Overconfidence can lead to frustration and failure since goals are rarely met. Accurate calibration can enhance motivation by affecting the attribution process. Poorly calibrated learners are subject to the fundamental attribution error, attributing their successes to competency and their failures to bad luck. A well-calibrated learner, in contrast, will make appropriate attributions for success and failure and be able to make effective use of feedback. Research on child chess experts (Horgan, 1990) suggests that unusually accurate calibration is associated with precocious expertise. Being able to set appropriate goals, having a clear sense of how much effort is necessary to reach those goals, being objective about one's own performance, attributing success or failure to appropriate reasons, focusing feedback to specific areas of weaknesses, successfully managing time so that one does not overprepare in areas already mastered while not preparing enough in weak areas--all of these activities result from accurate calibration and may be keys to unusually effective learning.

Another implication has to do with time management. Many successful students fail later as graduate students or in careers because of underconfidence--they have consistently overstudied (and been rewarded for this with good grades). When time demands increase, they are unable to cope.

## References


Glenberg, A.,& Epstein, W. (1987). Inexpert calibration. Memory & Cognition.

Horgan, D. (1990). Calibration in an expert domain: Child chess players' predictions of success on and off the chess board. International Symposium on the Psychology of Skilled Chess, Helsinki, June 1990.

Lichtenstein, S., Fischhoff, B. (1977). Do those who know more also know more about how much they know? Organizational Behavior and Human Performance; 20(2) 159-183.

Lichtenstein, S., Fischhoff, B., & Phillips, L. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky, (Eds). Judgement under uncertainty: Heuristics and biases. New York: Cambridge University Press.

Mabe, P., & Wells, S. (1982). Validity of self-evaluation of ability: a review and meta-analysis. Journal of Applied Psychology, 67, 230-296.